

CODERHOUSE

MODELO PREDICTIVO VALOR VIVIENDA AUSTIN - TX - EEUU

DATA SCIENCE

SEBASTIÁN CAMPOS V.





CONTEXTO E INFORMACIÓN

El mercado inmobiliario de Austin se ha destacado como uno de los más dinámicos durante el año 2021, y este listado proporciona una visión de cómo ha evolucionado en los últimos dos años. Los datos recopilados fueron obtenidos por un usuario de Kaggle, quien los utilizó en su proyecto final de análisis de datos en la Universidad Northwestern de Estados Unidos.

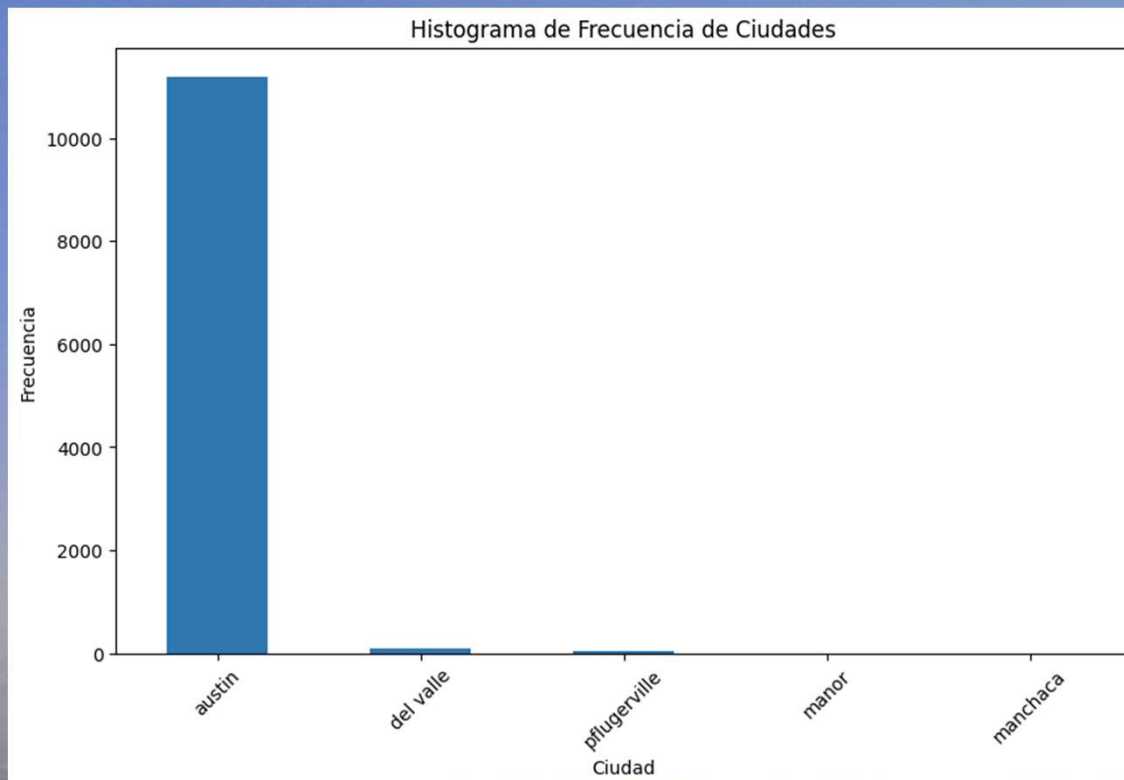


IMPORTANCIA

La importancia de llevar a cabo un análisis y modelo predictivo de los valores de inmuebles radica en la capacidad de estimar el precio de una posesión tan significativa y relevante como lo son nuestros hogares. Los inmuebles representan una de las compras más importantes que algunas familias realizan en su vida. Además, también son una fuente de ingresos rentable para muchas personas que se dedican a la inversión.

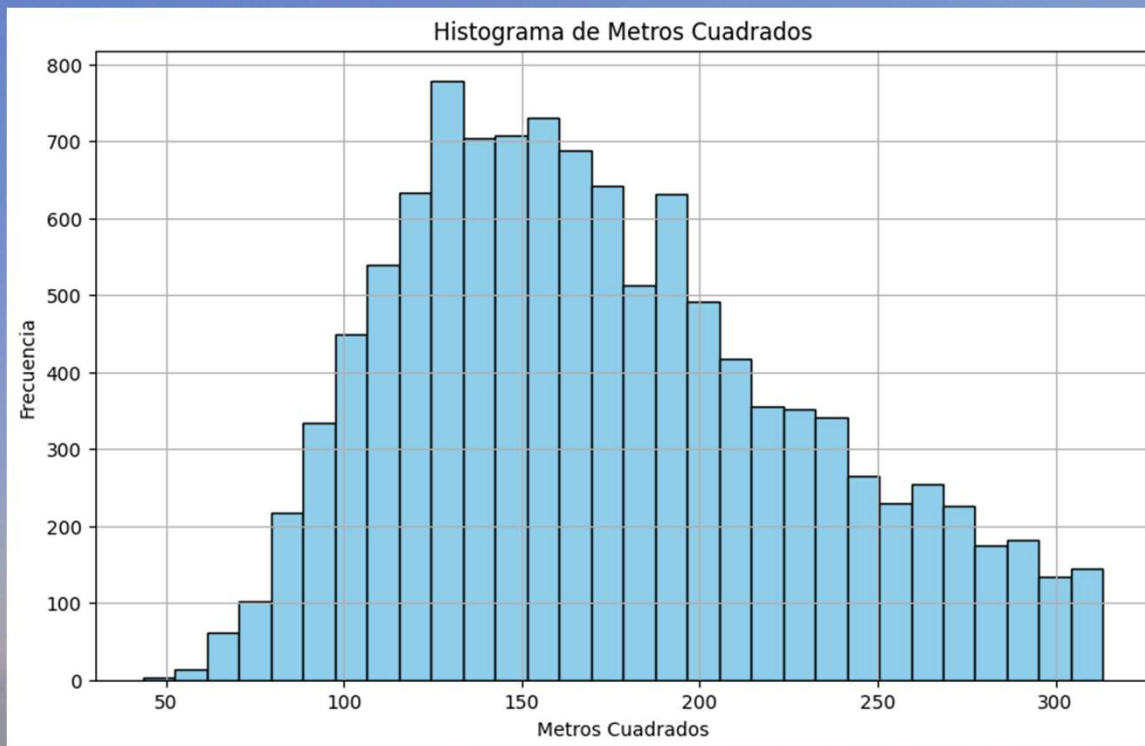


Análisis exploratorio



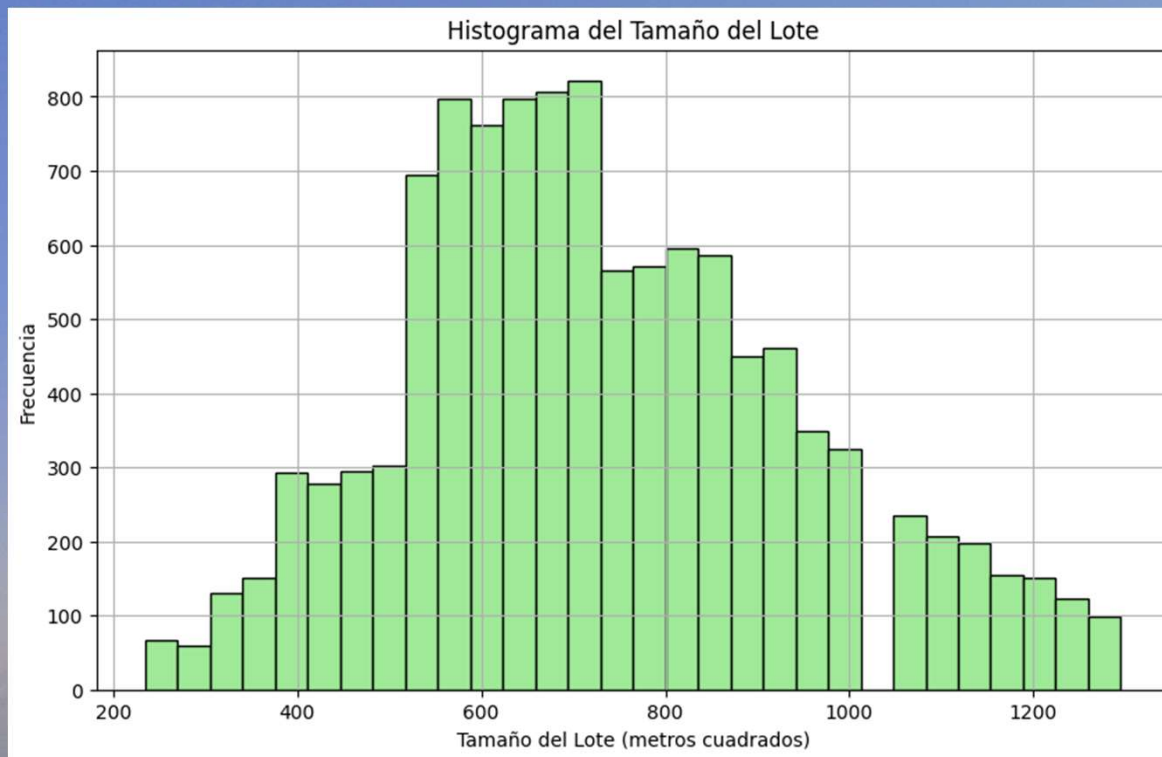
Se observa que la ciudad que estudia el proyecto de DS es principalmente de la ciudad austin en el estado de TX

Análisis exploratorio



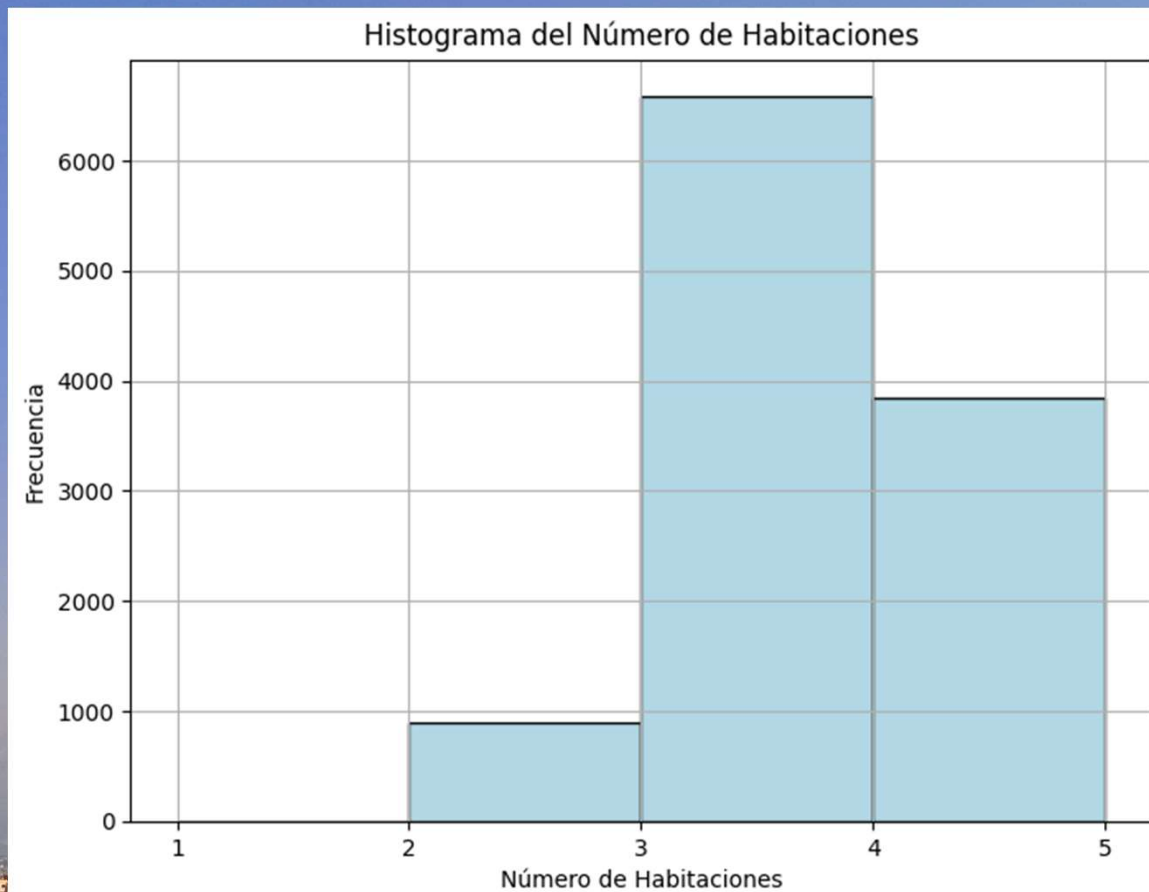
Se observa que la cantidad de m2 se concentra de 100 a 200 m2 construidos

Análisis exploratorio



Se observa que la cantidad de m2 de terreno es de 500 a 900 app

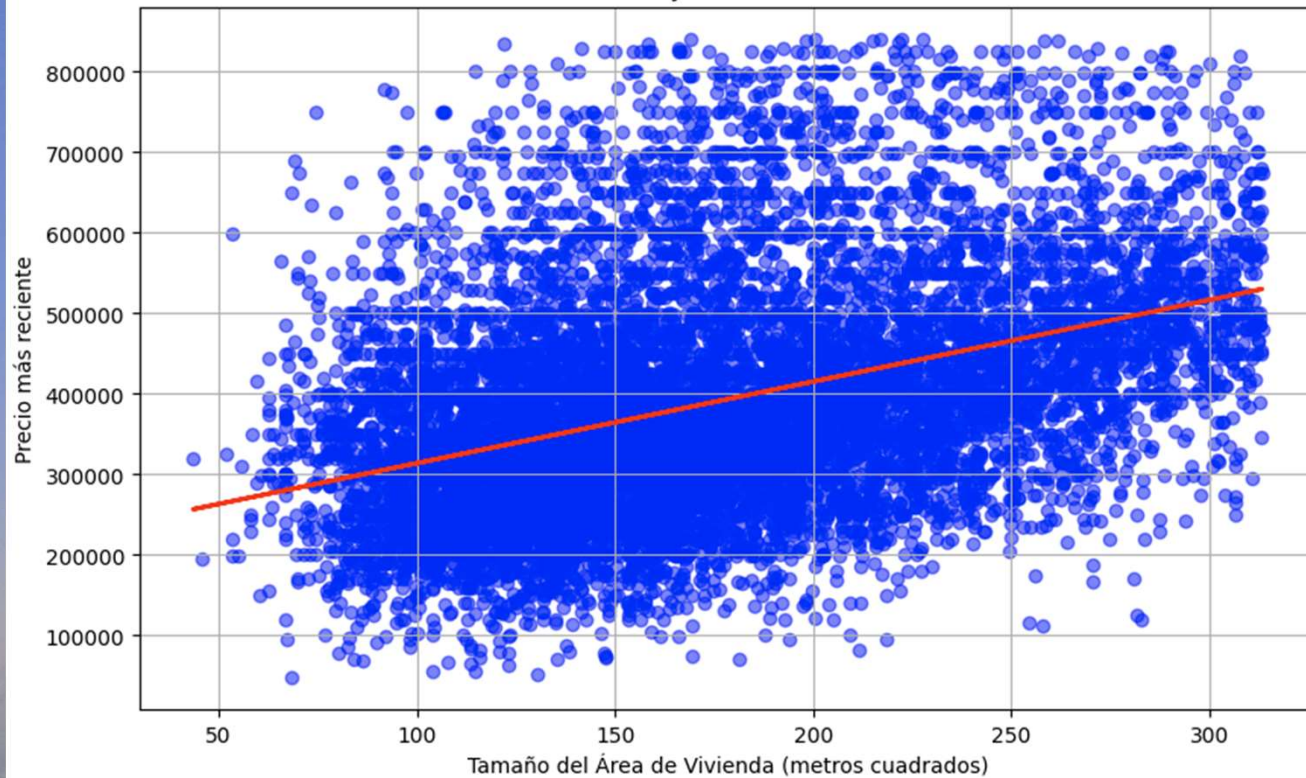
Analisis exploratorio



Se observa que la cantidad de dormitorios es principalmente de 3 a 4.

Análisis exploratorio

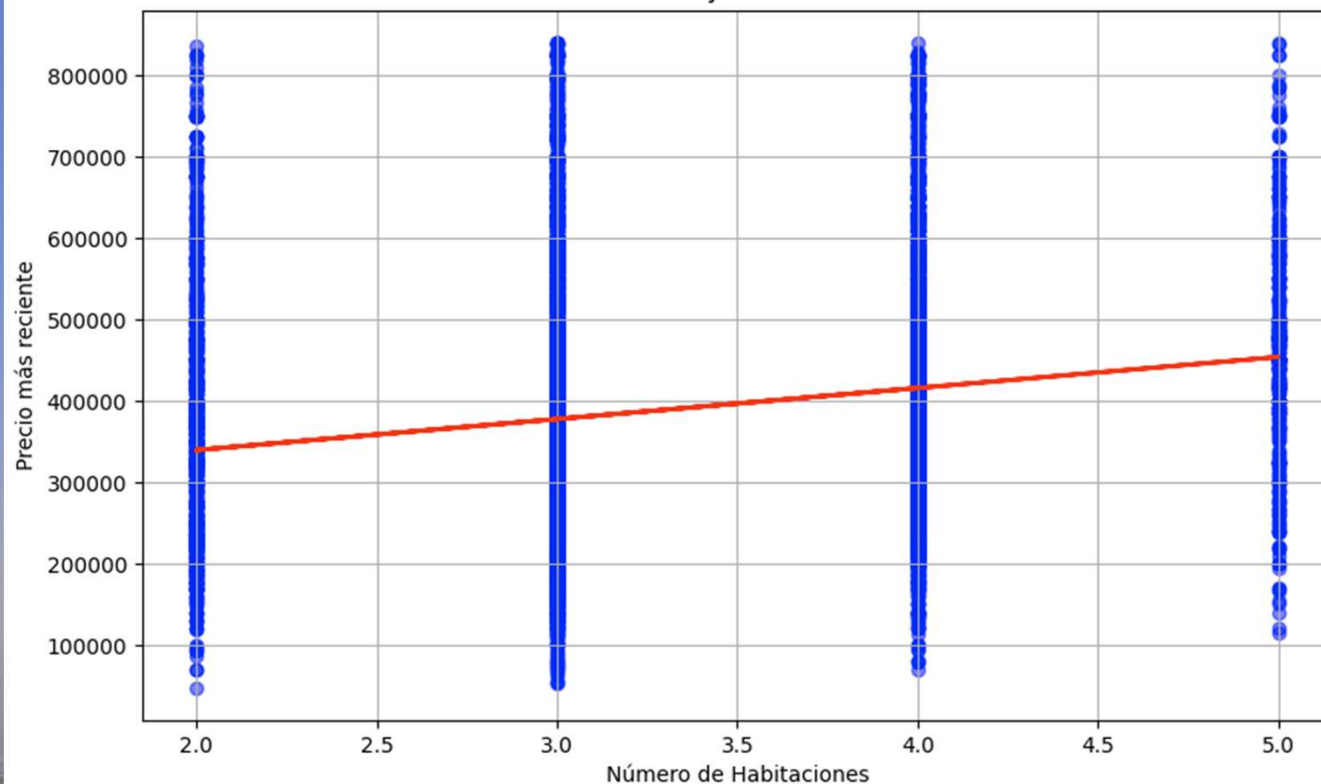
Relación entre Precio y Tamaño del Área de Vivienda



Se observa que la tendencia no es tan marcada pero si que a mayor cantidad de m2 construidos aumenta el precio.

Análisis exploratorio

Relación entre Precio y Número de Habitaciones



Se observa que la tendencia no es tan marcada pero si que a mayor cantidad de m2 construidos aumenta el precio.

INSIGHTS

Solo se está evaluando el mercado de austin



Estamos enfocados únicamente en evaluar el mercado de Austin y, según nuestros análisis exploratorios, hemos observado que, en promedio, las casas tienen entre 2 y 4 habitaciones, con una superficie construida de 170m² y un terreno de 700m².

Se trata de una ciudad bastante homogénea



Dado que principalmente se trata de viviendas unifamiliares, podemos notar que la ciudad exhibe una homogeneidad significativa. Esto posiblemente se refleje en una correlación menos pronunciada entre las variables y nuestro objetivo.

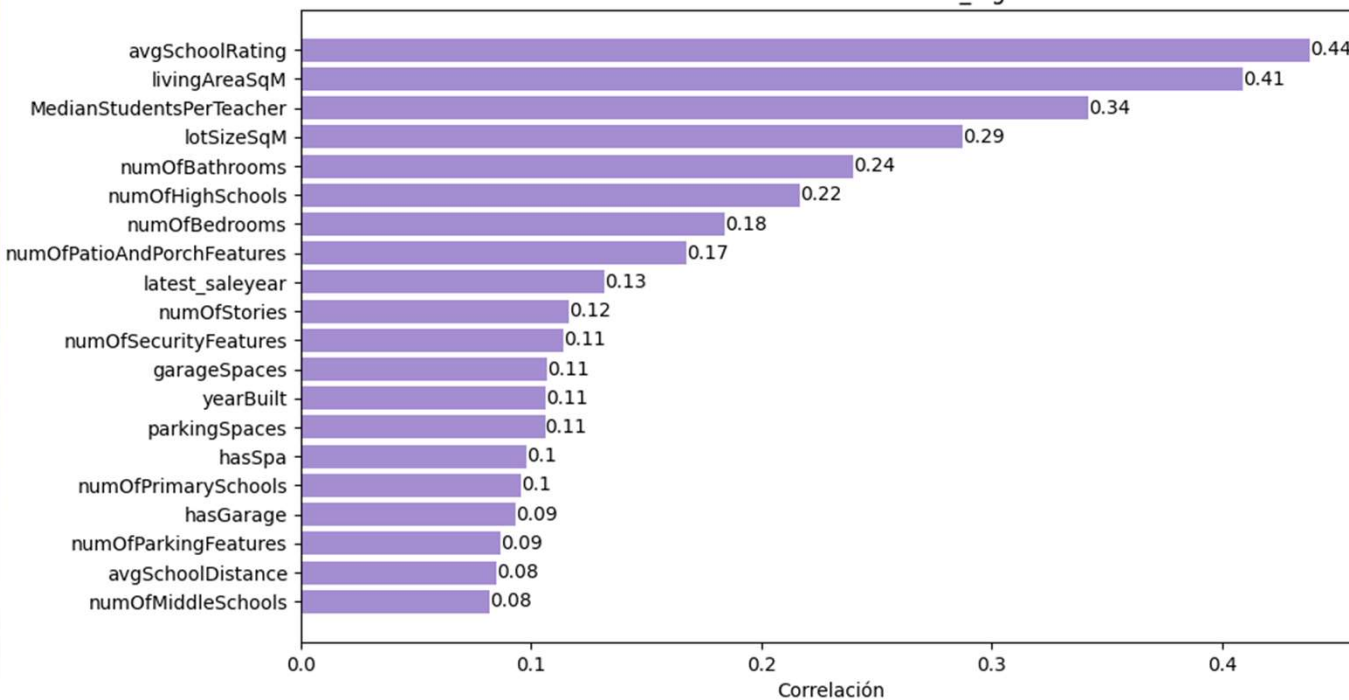
Otras cosas a valorar



Podemos inferir que, debido a la similitud de los datos, es probable que existan algunos detalles únicos que generen mayor relevancia que simplemente la superficie construida y del terreno. Estos detalles podrían incluir aspectos como los materiales de construcción, la cercanía con servicios, la presencia de características de lujo, entre otros.

INSIGHTS

Correlación con latestPrice_log



Correlación



Como se observó en los insights, una de las variables que muestra una correlación más fuerte es la proximidad a centros educativos. Posteriormente, las variables de área construida, que generalmente se considera importante, y la del terreno ocupan el segundo y cuarto lugar, respectivamente. Podemos concluir que el entorno educativo de la zona tiene un papel significativo en la definición del precio.



MODELO APLICADO

En el análisis de diversos modelos de regresión ha encontrado que el CatBoost Regressor es el que mejor explica el dataset. Con un Mean Absolute Error (MAE) aproximado de 0.1475, un Mean Squared Error (MSE) aproximado de 0.0458 y un R-squared de aproximadamente 0.6739, el CatBoost Regressor logra explicar alrededor del 67.39% de la variabilidad en los precios de las viviendas. Si bien el modelo no llega a explicar de mejor manera nuestros datos es importante reconocer que ninguna de las variables explicativas muestra una correlación muy alta. Este hallazgo sugiere que factores no cuantificables, como preferencias personales y características de lujo, también influyen significativamente en las decisiones de compra de propiedades.

GRACIAS

